Document: Frontier/Squid requirements for AGIS

Date: March 25, 2011

Change log

Date	Author	Description
03/11/11	Alexey Anisenkov	Initial requirements collected from email discussions
03/16/11	John DeStefano	Updates and additions
03/25/11	Dave Dykstra	Details, corrections, and clarifications

Frontier/Squid in AGIS. Functional description

Purpose

The purpose of this document is to provide detailed list of the functionality need to be supported by AGIS to store, retrieve and manage Frontier/Squid information. It also describes basic entries of Squid/Frontier data from functional point of view.

Description

Frontier servers. Some (not all) Tier-1s have them. Each server can consist of one or more physical machines, with an optional round-robin alias pointing at them. Jobs need to see both the round-robin alias and the individual machines behind them (the latter in case of fail-over), but monitoring needs to see only individual machines, so a distinction must be made somehow.

Squids. Almost all (but not all) sites have them. A site can have one or more Squids. If there is more than one Squid there is often a round-robin alias for the service name but just as with the frontier servers, jobs have to be able to failover to the individual machines and monitoring also needs to see the individual machines. Instead of a round-robin aliases, sites may instead choose to have the client do load-balancing between their Squids. Also, sites may have different, private net, addresses for internal access by jobs than the public addresses for external access by monitoring or by job fail-over from other sites.

Frontier servers also have one Squid instance built into each server, but these are deployed in a completely different mode than the site Squids, and they should be thought of and treated as part of the Frontier service. The address specified as the Frontier server is really the address of the Squid in this special mode.

Frontier server sites may optionally have backup-only Squids associated with them. These are used for fail-over in case a site's own Squids fail.

Squid listens on different ports for access by jobs and by monitoring. The default job port is 3128 and the default monitoring port is 5401, but both of them may be overridden.

Monitoring should be supplied for the real servers that comprise a site service (e.g., the

individual site URLs that exist behind a load balancer), but not for an alias (e.g., a load-balanced service). Sites may also have additional Squids that they want to have monitored but not accessed by any job. Conversely, sites may have some Squids that they want to be accessed by jobs but not be monitored. Monitoring is always done on a public address for a server, not on private addresses.

The names of one or more Frontier server sites must be associated with each Squid site. The names of one or more back up (failover) Squid sites can also be associated with a given site. The list of available Squids for a given site comes first from the private addresses on its own site, then from public addresses on backup Squid sites, and finally from backup Squids associated with the Frontier server sites

It is expected that the number of AGIS clients will be proportional to the number of sites in service, and not to the number of worker nodes at each site.

At present, AGIS will poll for data from external sources, with TiersOfAtlas as a primary data source, and other data coming from GOCDB, MyOSG, and BDII. Once the functionality and reliability of AGIS has been established, it is expected that AGIS itself will assume the role of the primary provider of this data.

Requirements

Client view

AGIS should provide the following data sets to client requests at a site:

- List of Frontier servers available for jobs (`serverurl`s), from the perspective of the site on which the client is running. A Frontier server definition consists of an HTTP URL, in the form of a host, colon, port number, slash, and servlet path. The order is important. A round-robin alias, if present, should be first. If there are more than 2 individual servers, the round-robin alias should appear twice. Next should be the individual servers. Following that should be the same items from backup Frontier server sites, in order.
- List of Squid servers available for jobs (`proxyurl`s) at the client site. A Squid server definition consists of an HTTP URL in the form of a host followed by a colon and port number. There also should be an indication of whether or not the client should do load balancing between all the Squid servers. The order is important, unless the client is doing the load balancing. The same rules for round-robin aliases apply as for Frontier servers.
- List of backup Squid servers available for jobs ('backupproxyurl's) from remote sites. Backups Squids may come from either Squid sites or Frontier server sites. Only external addresses are used on the remote sites. It is an error for remote Squid sites to specify client-based load balancing, because frontier_client cannot be configured to support that; sites listed as having backups must have a round-robin alias (or have only one Squid).

Monitoring view

AGIS should provide the following information to monitor requests:

- List of Frontier servers and squid servers that are to be monitored, in the form of an HTTP URL of the external name of the servers followed by a colon and port number. Only external addresses may be used

Server data

AGIS should provide the following data sets for each ATLAS site:

- List of Frontier server names provided by the site, each with a host name, job-access port, servlet path, and optional non-standard monitoring port. Only a small number of sites actually have Frontier servers so it might make more sense for this to be an entirely different type of site and not something that is available for every site to fill in.
- List of Squid server names provided by the site, each with a host name, optional non-standard job-access port, and optional non-standard monitoring port.
- When either a Frontier or Squid server is a round-robin alias (or hardware load-balancer) and there are more than 2 servers behind it, the round-robin alias should be listed twice.
- Flag that says whether or not the client should do load balancing between the site's own Squid servers.
- List of backup-only Squid servers, in the same form as the Squid servers list. These are only on Frontier server sites.
- List of backup Squid server site names for jobs to use.
- List of Frontier server site names for jobs to use. Any backup-only Squid servers supplied by those Frontier server sites should automatically be added to the end of the backup Squid server sites list when a client requests data.
- For each Squid or Frontier server definition, additional flags should be provided to determine whether the following conditions apply, all defaulted to “no”:
    - Should the server not be monitored? This should be “yes” when the server definition is of a round-robin alias, or if it it is on a private network.
    - Should the server not be available to local jobs? This should be “yes” if the server is is a public-only address (for monitoring or remote job access) or is for development/testing.
    - Should the server not be available to remote jobs? This should be “yes” if this site is listed as a backup site for some other site but the server is on a private network or otherwise shouldn't be used. It will never be use on Frontier servers.

API details

The AGIS API should include the following functions:

- list_frontier_configuration(name=[site_name]) – return information for local Frontier jobs
  - .squids – list of local Squid host URLs (will be used by 'proxyurl')
  - .frontiers – list of Frontier server URLs (will be used by 'serverurl')
  - .backup_squids – list backup Squid host URLs (will be used by 'backupproxyurl')
  - .client_balanced – flag that says whether or not client should do load balancing between local squids
- list_squid_configuration(name=[site_name]) – list only squid configuration, for applications other than Frontier that use squids
  - .squids – list of local Squid host URLs
- list_monitor_configuration(name=[site_name]) – list configuration for remote monitoring
  - .squids – list of monitor URLs for the site